What to Expect When You're Accessing: An Exploration of User Privacy Rights in People Search Websites

Kejsi Take
New York University
New York, NY, USA

Jordyn Young
University of Michigan
Ann Arbor, MI, USA

Rasika Bhalerao
Northeastern University
San Francisco, CA, USA

Kevin Gallagher
NOVA LINCS, NOVA School of Science and Technology
Lisbon, Portugal

Andrea Forte
University of Michigan
Ann Arbor, MI, USA

Damon McCoy
New York University
New York, NY, USA

Rachel Greenstadt
New York University
New York, NY, USA

ABSTRACT

People Search Websites, a category of data brokers, collect, catalog, monetize and often publicly display individuals' personally identifiable information (PII). We present a study of user privacy rights in 20 such websites assessing the usability of data access and data removal mechanisms. We combine insights from these two processes to determine connections between sites, such as shared access mechanisms or removal effects.

We find that data access requests are mostly unsuccessful. Instead, sites cite a variety of legal exceptions or misinterpret the nature of the requests. By purchasing reports, we find that only one set of connected sites provided access to the same report they sell to customers. We leverage a multiple step removal process to investigate removal effects between suspected connected sites. In general, data removal is more streamlined than data access, but not very transparent; questions about the scope of removal and reappearance of information remain. Confirming and expanding the connections observed in prior phases, we find that four main groups are behind 14 of the sites studied, indicating the need to further catalog these connections to simplify removal.

1 INTRODUCTION

Privacy regulation in the United States is based on the "notice and consent" paradigm – users are notified of the terms of use and can choose to accept the terms or refrain from using the service [38]. An objection of this paradigm posits that it is unlikely that these notices contain comprehensive information about how data is retained, analyzed and distributed now or in the future [31, 38, 40]. Naturally, this results in unexpected flows of personal information [31]. One of the main actors behind extensive data collection, aggregation and monetization are data brokers [12]. While these companies sell and buy various types of data, People Search Websites (PSW), a unique category of data brokers, create and publicly display individuals' profiles or "dossiers" without their permission. These profiles include Personally Identifiable Information (PII), such as physical addresses, phone numbers, potential family members, and legal records, exposing users to many risks [36, 44].

Indeed, multiple guides recommend that people who wish to limit the amount of their publicly available information request that their information be removed from these sites [8, 22, 25]. Additionally, prior work has found that individuals often seek removal from People Search Websites due to concerns related to the potential misuse of PII by harassers or social engineering attackers [44]. This prior work was based on participants' recollections and one of its key findings was that users were concerned about the opacity and efficacy of the removal process. In this work, we systematically analyze People Search Websites' data access and removal mechanisms to directly address this research gap and better inform researchers, users, and policy makers about data access and removal in practice.

We collect and analyze researchers' own observations of the access and removal processes to understand these sites. This method allows us to systematically study the sites by investigating researchers' experiences. Our different backgrounds and physical locations allow us to collect data relevant to understanding the usability of access and removal as enacted under both the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA), privacy regulations that codify individuals' right of access, opt-out and deletion. Further, due to historical evidence of multiple mergers and acquisitions among PSWs [13, 26], we wanted to study their effects on the removal process. Many sites are owned by the same companies, but is not clear how removal requests are treated across such connected sites. To answer this question, we designed a two-step removal process in which we first sought removal from a set of suspected connected sites and then observed removal effects in all of the remaining sites, before requesting removal from the rest.

In terms of data access, we found that most sites are reluctant to provide access to an individual's data upon request, and often offer alternatives such as removal (without specifying what exactly there is to remove) or recommendations to self-search. Other times, requests are rejected without alternatives. Sites' responses shed

Proceedings on Privacy Enhancing Technologies 2024(4), 311–326

https://doi.org/10.56553/popets-2024-0118

light on the reasons why this might be happening, indicating that under the CCPA, the information collected by these sites is "public" and as such is excluded from the legislation. Similarly, we provide examples of potential GDPR misinterpretations resulting in additional data access rejections. We also obtain access to paid reports and provide details about the types of information they contain and the capabilities they unlock for malicious parties.

Further, the two-step removal process enabled us to observe that removal on certain sites leads to removal from other connected sites. However, these sites provide little transparency or indication about the scope of removal. We found that issues with identity verification complicate removal.

We combine insights from the data access, data removal, and paid reports to provide insights on the monetization ecosystem of People Search Websites. We find that there are three main types; *fee sites* that sell reports about various search targets, *ad sites* that advertise them and *hybrid sites* which do both. By observing common access mechanisms and removal effects, we find that 14 of the sites are connected to four groups. Additionally, we find that *only one of these groups studied (BeenVerified) responded to access requests with the same report given to their paying customers.*

The existence of multiple People Search Websites makes privacy self-management complicated, but our insights can be used to develop more efficient removal strategies. More specifically we find that (1) systematic measurement can provide insight to these sites practices and policies around data access and removal, (2) access requests are mostly unsuccessful and existing regulations are not effective, and (3) better understanding connections between sites can simplify and improve removal. We also discuss policy implications and provide suggestions for improving usability of users' privacy rights.

2 BACKGROUND

In this section we present an overview of data brokers and People Search Websites, and relevant regulations. This is not intended to be a comprehensive evaluation of the law, but rather key points that are relevant to contextualize our findings.

2.1 Data Brokers and People Search Websites

Data brokers are companies whose primary business is collecting personal information about consumers from various sources and aggregating, analyzing, and sharing that information, or information derived from it [12]. Recent work has examined major data brokers and the highly sensitive data they hold on U.S. individuals and found that these records include tens of thousands of sub-attributes ranging from personal preferences [36] to mental health data [21]. People Search Websites are also a type of data broker. Prior work defines them as sites that allow internet users to search for information on an individual by entering their name [36, 44]. Services that remove information from these sites have emerged [1] [2]. Although, prior work found that in order for them to be effective one must keep continuously paying [44].

[1] https://joindeleteme.com/
[2] https://privacyduck.square.site/

2.2 Relevant Regulations

Data privacy and security for European Union (EU) users is regulated by the GDPR (General Data Protection Regulation). This law, while drafted and passed by the EU, imposes obligations into organizations anywhere, so long they target or collect data related to people in the EU [51]. Relevant rights covered by the GDPR include "the right of access"(Art. 15) [5] and "the right to be forgotten" (Art. 17) [6] that enable subjects the right to obtain a copy of or delete their personal data.

While the U.S. does not have a general privacy law, recently some state-level privacy laws have been introduced. The most established among them, the CCPA provides California residents the right to know (Cal. Civ. Code § 1798.110) [1] and right to delete (Cal. Civ. Code § 1798.105) [2], similar to the GDPR's right of access and erasure. A few states have privacy laws dedicated to data brokers [3, 4], but at the time of the writing none of them delineates a requirement to allow individuals to opt out.

2.3 Personal Information Definitions

While CCPA is widely regarded as GDPR's equivalent in the United States, there exist some fundamental differences between the two regulations [48]. Most relevant to our work are the different definitions of personal information. In this definition CCPA excludes publicly available information (i.e. government records) [30], while the GDPR does not.

California Privacy Rights Act (CPRA) [3], the most recent CCPA modification, extends the scope of publicly available information to specify "information that [...] is lawfully made available to the general public by the consumer...". This addition further expands the types of information not included in the definition of "personal information" [14]. In contrast, the GDPR does not have such exclusions, defining "personal data" as "any information relating to an identified or identifiable natural person ('data subject'); [...]such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person." This divergence is essential to contextualizing our findings because public government records are a common source of information for data brokers and People Search Websites [12]. In section 5 we find that the CCPA limitations outlined in this section are used to justify the rejection of access requests.

2.4 Misuse of Personal Information

Recent technological advances have enabled a wider reach to personally identifiable information. Nowadays, attackers can easily use online sources, such as PSWs, to obtain targets' PII. Indeed, the role of personal identifiable information in online harassment campaigns, such as doxing or swatting, has been extensively documented [7, 15, 39, 45]. Research investigating online infidelity forums has found that perpetrators may use PSWs for reverse lookups to identify unknown numbers from a partner's text or call records [46]. Further, recent work investigating security and privacy advice confirms that minimizing publicly available information is a top priority for potential targets of online harassment [49], which includes taking steps to remove information from PSWs. Our work provides

[3] https://thecpra.org/

additional support for this recommendation with evidence that PSWs facilitate access to sensitive information. We also contribute insights into the removal processes and strategies for making them more effective.

3 RELATED WORK

Prior research on People Search Websites is limited. Researchers have discussed these sites as threats to national security [37]. Recent work investigated reasons why participants seek removal from PSWs and analyzed these experiences [44]. However, research has yet to systematically explore the data access and removal processes and their outcomes. To fill these gaps, in this work we aim to (1) understand how PSW data access works, (2) provide an overview of removal mechanisms and (3) discover connections between the different sites. In the rest of this section, we discuss relevant prior work investigating user privacy rights.

3.1 Data Access

While a right to data access is included in both CCPA and GDPR, researchers have documented problems with companies' processes to implement these rights. Prior work studying data access requests has documented issues with delayed or missing responses to data access requests, incomplete data access outcomes, and challenges with identity verification [9, 23, 42, 47]. To our knowledge, no study has systematically explored and documented the challenges that emerge when seeking access to one's own data in PSWs specifically. Unlike user self-generated content on social media, PSWs publicly display profiles that have not been created by individuals themselves. This is particularly problematic for targets of online harassment, as they cannot engage in privacy-protecting behaviors (e.g., changing leaked contact information) without knowing what information attackers might have accessed through PSWs.

Besides studying the privacy rights mechanisms, our intent was to use reports obtained through data access requests to better understand the sites data collection practices. Data access requests have been used as a method to enhance transparency when external access to data is limited [10, 50, 52]. However different from sites in prior work, compiling and selling user profiles' is the main source of profit for PSWs and therefore they might be disincentivised to provide free access.

3.2 Data Removal

Prior research has indicated that data deletion mechanisms in non-PSW contexts have usability issues [19, 20]. On social media, for example, deletion interfaces often use dark patterns, including confusing language [35]. Mobile apps users who try to delete their accounts also find the deletion process complex, with the most frequently reported issue being difficulties in locating the privacy rights mechanism [27].

Recent work studied GDPR compliance of a set of online services by first requesting erasure of their data and later sending data access requests to find out what and if data remains [34]. This work explores a context where authors themselves created accounts that they later attempted to delete. PSWs are unique in that they create public profiles without users' permission, making it more important to study user privacy rights, as the only ways to reinstate user agency. A recent study explored people's experiences removing their information from PSWs specifically [44]. They found that participants not only reported dark patterns when attempting information removal but also noted that information could subsequently reappear. While [44] was limited to participants' experiences with data removal, our study systematically explores what both the data access and the removal process look like across sites and evaluates the outcome of these processes.

Further, both [35] and [27] found that usability hurdles can cause users to abandon the deletion process, which can be particularly problematic – or simply not an option – for users seeking removal from PSWs due to immediate safety concerns. In this work, we investigate connections between the sites that can be leveraged to simplify the removal process.

4 METHODOLOGY

In this section, we discuss ethical considerations, steps we took to thoroughly consider them in our study design and website selection methodology. We also describe the phases of the study, which are summarized in Figure 1.

4.1 Ethics

During the study design phase, we had multiple discussions about what methodology could be used to best answer our research questions. While it is typical for usable security and privacy research to recruit and study participants' experiences through lab studies or research interviews [17, 19], previous work on People Search Websites indicated that they contain sensitive information and documented participants' concerns about the handling of their personal data by such sites [44]. To avoid soliciting such sensitive data from participants and to make sure we document every step of the process thoroughly, we decided to borrow data collection techniques from auto-ethnography [11] studying the experiences of the research team members themselves. We took measures to protect the researchers participating in the study, including ensuring all were informed about the potential risks and had the agency to opt-out from any site interactions they felt uncomfortable with or withdraw from the study at any point. The researchers created burner study email addresses to avoid using their personal ones. In addition, interacting with the sites was not a requirement to participating in the study. This study was deemed not human subjects research by the Institutional Review Boards (IRB) at New York University.

Further, the research team also considered the burden on the websites themselves. While some burden was necessary to answer our research questions, we aimed to minimize it by only emailing customer service when automated data access and removal mechanisms failed or were nonexistent. If they indicated the process was outside of their scope, we did not push them. In doing so, we avoided asking customer service to perform any tasks outside of their standard responsibilities in these areas.

4.2 Study Design

To provide a more systematic overview of the information provided by these sites and individuals' privacy rights mechanisms under different privacy jurisdictions we followed a methods similar to

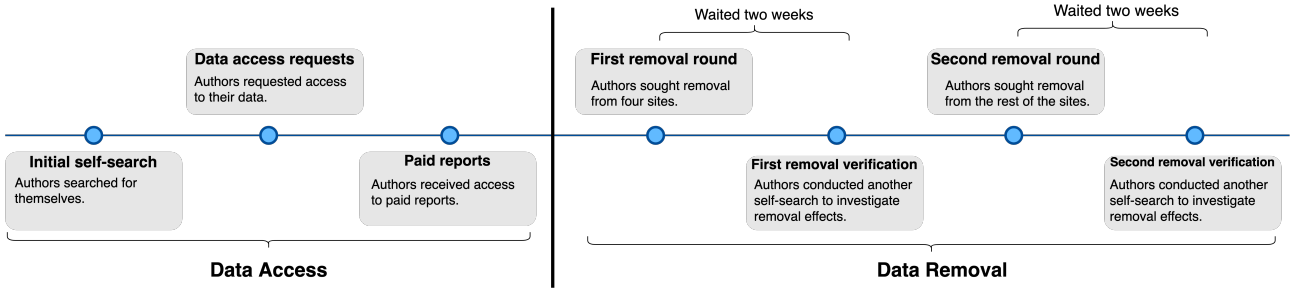

Figure 1: Overview of the data collection steps: The first three steps are focused on investigating data access and the last four in the data removal.

the ones in prior research [20, 34], analyzing researchers' experiences instead of recruiting participants. Differently from Habib et al. [20], Rupp et al. [34] who in order to study the sites in their datasets, created new accounts with fictional persona, People Search Websites had already accumulated information and created profiles without our consent. For this purpose, we developed analysis templates to track data access and data removal across multiple People Search Websites. This goal is similar to Spiller [42], who used subject access requests of his own data to study the handling of urban CCTV digital data. Lastly, privacy journalists used authors' information to study the ecosystem of sites spreading slander [24].

4.2.1 Positionality. In order to study privacy rights mechanisms under different jurisdictions, we assembled a team of researchers with different backgrounds. Among them was a California resident, an EU resident (previously based in the U.S.), and two non-California residents, based in the U.S. Two of the researchers have fairly common names. While we would not normally include information about the researchers, here we find it necessary to do so with their permission, as residency status determines the jurisdictions of different privacy laws. To respect the authors' privacy, throughout the paper we only mention the state of the residence when relevant due to privacy regulations.

4.2.2 Study Overview. For the purpose of our study, we define access as understanding the extent and types of personal data that these sites contain and offer to their customers as part of their services. Further, we define removal requests as mechanisms that allow users to request that these sites stop displaying their personal information. We note that defined this way, "removal" is different from data deletion, a privacy right that enables users to delete accounts or information related to their accounts. Researchers were instructed to use their browsers private mode for searching themselves.

4.2.3 Part 1: Access. In order to gain a comprehensive understanding of the personal data that People Search Websites have on users, we approach this topic from these vantage points: initial self-search, to understand the data that is available to a non paying website visitor; data-access request, to understand the information that these sites collect about subjects and paid access, to understand the information that is available to a paying website visitor.

4.2.4 Part 2: Removal. Given that the information that these sites display (i.e. physical address, phone number) is highly sensitive, we also studied the usability of profile removal that these sites offer. In the access phase, we observed that multiple sites partnered with or redirected with each other. Therefore, as we discuss in 4.4, we organized the data collection, to investigate if the removal from one site also removed information from connected sites.

4.3 Website Sample

To select the subset of sites to analyze, we started from an initial list of People Search Websites and ranked them according to Tranco's rankings of most popular sites [32]. We programmatically compared the Tranco ranking with the sites included in the most comprehensive data removal list we found [4], to extract the most visited People Search Websites. However, this list also contained general data brokers and other data collection companies. Our goal was to study the People Search Websites, and allow searching an individual by name and publicly display their PII. We excluded sites that require an account for two reasons. First, making an account requires giving the site information about ourselves (e.g., email address), but we wanted to see what information PSWs have without any input from a target. Second, prior work [19, 20, 34] only studies sites that require account creation, so we address a gap in the literature to understand sites that allow data access without requiring an account. This way, we ended up with an initial set of 20 sites. We removed a site that redirected to Intelius after searching, but kept AnyWho to help understand how shared site ownerships affect removal. To replace the removed site and investigate the same question in regards to Whitepages, we added 411, a site we found to share the same privacy policy, bringing the total number of sites to 20. The final list of sites is displayed in Figure 2.

To investigate the generalizability of our findings we conducted a cursory search to find how many of these websites turn up in the first 50 results on Google with two different search terms for two of the authors("[first, last name] address" and "[first, last name] phone number"). Among these results we found 14 unique sites that fit our definition of PSWs (sites that allow searching an individual by name and publicly display their PII), ten of each were in our dataset and four others linked to sites in our dataset. We conducted the search in an incognito tab. However, this measure reduces but does not completely eliminate search results personalization. Further,

[4] https://inteltechniques.com/workbook.html

we compared sites in our sample with those included in a removal guide from Consumer Reports [18]. Our dataset contains 11 out of 14 sites that the article suggests prioritizing if seeking removal.

4.4 Data Collection

In this section, we explain the data collection process in more detail.

4.4.1 Setup. The first author performed an initial exploration of the sites and compiled comprehensive templates for each phase to enable standardized recording of the researchers' experiences of interacting with these sites. We summarize the instructions in Section 4.4.2. The first author also set up accounts and payments for each site and downloaded reports for themselves and the other researchers. For the phases that included interacting with the websites to exercise one's privacy rights, and to generate instructions for the rest of the researchers, the first author explored the sites, focusing on the privacy policy and privacy rights options in the website footer. However, the instructions and prompts for the semi-structured note-taking were discussed with other authors too, and modified as necessary.

4.4.2 Instructions . Here we summarize the set of instructions given to researchers for data collection. We also note that instructions for each phase included recommendations for recording each listing we found, each step we took during the search process, and notes for any other observations. Researchers were asked to:

- *Initial self-search.* Search for themselves in each of the sites in the subset and record the information found and its accuracy.
- *Data access requests.* Use the mechanisms set in place to request access to one's data, seek to obtain access to your data report. Make notes of all steps taken to do so, including forms completed, emails exchanged with the site, potential follow-ups, and any reports and replies received.
- *Paid reports.* Use the paid reports that the first author saved, record all types of information included and its accuracy.
- *First removal round.* Follow the instructions provided by the first author to attempt to request removal from four main sites (Intelius, BeenVerified, Peoplefinders and Whitepages) and not the rest. We choose these sites due to suspecting they were connected to other sites, either through shared privacy policies or advertisements. We describe these connections in Appendix A.
- *First removal verification.* Check if the removal was successful on the list of four sites. Check if the removal of information has resulted in removal from any of the other sites too.
- *Second removal round.* Attempt to request removal from the second set of sites.
- *Second removal verification.* Verify that the information was removed from the second set of sites.

To further ensure the thoroughness of data collection, researchers involved in the data collection met regularly. In these meetings, researchers raised issues encountered and interpreted results. We also discussed modifications to our course of action as necessary, for example we decided when to follow up with sites that did not respond to our requests. In addition to the structured data collection, researchers also took semi-structured field notes, also used in previous research studies [16, 29].

4.5 Limitations

First, our main goal is to understand People Search Websites. As such, our study does not attempt to provide any insights on how individuals find and exercise privacy choices. While we describe the difficulties encountered while trying to exercise these choices, the main goal is to analyze the outcomes of these actions systematically. Further, the researchers exercising their privacy rights have higher levels of education than average, and are also familiar with privacy research and best practices. As such, the typical internet user might find this process even more difficult when trying to exercise their privacy rights given the obstacles we encountered. However, the goal of our study was to understand the sites, not how individuals interact with them and a comprehensive user study is outside of the scope of this paper. Future work could recruit a diverse participant pool to explore users mental models and experiences of interacting with these sites.

Further by exploring the privacy mechanisms ourselves, we have visibility into all of the steps followed and therefore gain a systematic overview of People Search Websites. Had we conducted this study with participants, it would not have been clear if inconsistent observations were due to the steps followed or an indication of differential treatment. Second, we focus on a limited sample of People Search Websites. A cursory web search reveals that many more exist. While we cannot make any claims that our findings generalize to all the sites, our findings indicate that it is likely that at least some of these sites are impacted by removal in the main People Search Websites. Further, our methodology of selecting sites that rank higher in the Tranco list means that we included sites that are most likely to be visited when someone is seeking to use or misuse a People Search Website. Future work could explore connections between a larger set of sites.

5 RESULTS

In this section we outline our findings, organizing them according to the different study phases.

5.1 Initial Self-Search

During the initial self-search we found that all sites included at least one of the researchers and nine sites included all four. We also observed that while the researcher from the EU was listed, all information was from their time in the United States. We include the number of researchers found on each site in table 1.

EU visitors experience IP Blocking. We found that nine of the sites attempt to block EU IPs. To overcome this, the researcher used a popular VPN service to access the services. However, even when using this service, the researcher was unable to access MyLife and 411. While this might be an attempt to avoid GDPR compliance, it also makes it even more difficult for EU residents, that might be listed in these sites to exercise their privacy rights.

Different sites have different monetization strategies. During the process of searching for ourselves, we observed three types of sites based on their monetization strategies. The first type, we call *fee sites*, display little to no personal information on search subjects and hide access to more information behind a paywall.

Sites/**Group**	**Tranco's Ranking**	**Type**	**Nr. Researchers**	**Information before payment** Address	Phone	Email	**Information after payment** Address	Phone	Email
Whitepages									
whitepages.com	5026	fee	3/4	partial	no	no	full	yes	yes*◇
411.com	76766	ad	2/4	full	partial*	no	n/a	n/a	n/a
BeenVerified									
beenverified.com	8996	fee	4/4	no	no	no	full	yes*◇	yes*◇
peoplelooker.com	47939	fee	4/4	no	no	no	full	yes*◇	yes*◇
PeopleConnect									
intelius.com	18423	fee	3/4	partial	no	no	full	yes*◇	yes*◇
ussearch.com	43434	fee	3/4	partial	no	no	full	yes*◇	yes*◇
instantcheckmate.com	16918	fee	3/4	partial	no	no	full	yes*◇	yes*◇
truthfinder.com	17198	fee	3/4	partial	no	no	full	yes*◇	yes*◇
anywho.com	42102	ad	3/4	full	no	no	n/a	n/a	n/a
yellowpages.com †	2624	ad	2/4	full	yes◇	no	n/a	n/a	n/a
PeopleFinders									
peoplefinders.com	25663	fee	4/4	partial	no	no	full	yes	yes*◇
searchpeoplefree.com	43839	ad	4/4	full	yes*◇	no	n/a	n/a	n/a
truepeoplesearch.com	10470	ad	4/4	full	yes*◇	no	n/a	n/a	n/a
fastpeoplesearch.com	19052	ad	4/4	full	yes*◇	no	n/a	n/a	n/a
Independent Sites									
clustrmaps.com	5617	ad	1/4	full	yes*◇	no	n/a	n/a	n/a
spokeo.com	7151	fee	3/4	partial	partial*◇	no	full	yes*◇	-
mylife.com	12008	fee	4/4	partial	no	no	full	yes*	yes*
nuwber.com	16155	hybrid	3/4	full*	yes*◇	no	full	yes*◇	yes*◇
peekyou.com	21321	ad	4/4	partial*	partial*	no	n/a	n/a	n/a
radaris.com	33904	fee	3/4	partial	no	no	full	yes ◇	yes

Table 1: Summary of connections between sites, and information visible before and after payment. † indicates unclear group membership. Yellow Pages has no shared mechanisms with other PeopleConnect sites, but is affected by removal. * indicates observation only in some of the researchers profiles. ◇ indicates that the information for at least one researcher was incorrect. *Fee sites* displayed partial addresses before payment, and complete ones after payment.

This paywall prompts the user to subscribe to the site to be able to access search results. The second type, *ad sites*, display information without any purchase options, and contain advertisements for the *fee sites*, driving traffic in their direction. The third type, that we call *hybrid sites*, are a combination of the first two types, both selling access and advertising other sites. The type of each site is described in Table 1. It's important to note that the advertisements are embedded in the site, not through third party ad services. This is observed through obtaining similar ads through multiple visits. Most of these ads consist of affiliate marketing, a form of advertising where the merchant hosting the affiliate links receives a commission on each successful purchase.

However, the advertisements are not enough to draw definite conclusions about the connections or shared ownership between sites. For example, at the time of writing we observed that SearchPeopleFree advertised BeenVerified, PeopleFinders and Instant Checkmate which we later observe are part of different groups.

***Fee sites* hide information after a paywall.** While in *ad* sites we can see our personal information, *fee* sites and *hybrid* sites, typically contain partial to no information and then require payment for the complete report. As visible in Table 1, in most of the *ad* sites we can see complete addresses. Naturally, *fee* sites display little to no information about the search target. Among these, the most commonly displayed data points are: city, state, age, and related people, which we found to commonly include family members.

***Fee sites* collect searcher's information.** Two of the sites, BeenVerified and PeopleLooker do not display any information at all before payment, even after prompting users for their email and their first and last name. Once these data points are provided, it is revealed that search results are hidden after a paywall. PeopleFinders and sites in the PeopleConnect group follow a similar technique. As indicated in Table 1 they display partial information beforehand.MyLife not only required searcher's email to conduct a search but also their birthday. Typically, e-commerce sites collect customer's contact information after they decide to purchase. The reversed order may be a way to form an extended advertising relationship with the potential customer; we received marketing emails from multiple *fee sites*.

One of the email addresses provided during this phase was added to the paid report. We found that one of the email addresses used in the search was later added to the report of one of the authors from MyLife that we later accessed through purchase. We are not sure why we observed this only in the case of one of the

researchers, instead of everyone, but this finding implies that by attempting to find out what information certain sites have about them, users may be exposing themselves to privacy risks. Additionally, it raises questions about potential misuse of data shared with these sites during access or removal. It makes it difficult for users to provide data requested during the the removal process when it is not known how the data provided as part of the process will be used. Lastly, this finding also reinforces the difficulties of studying these sites; recruiting participants might expose them to privacy risks.

Advertisements can be misleading. Prior work on People Search Websites has indicated that they often make exaggerated claims about the types of information that they offer as part of their services [44]. Our findings confirm this observation. For example, search results from Spokeo indicate that there were court records found under the names of two of the researchers. However, when we purchase the reports we find out that there are no court records associated with any of the addresses they are associated with. We observe a similar strategy in PeekYou, a site that includes ads suggesting that BeenVerified includes author's emails from certain domains, but paid reports do not include these emails. This observation confirms concerns about reputation damage and exaggerated claims discussed by participants in previous work [44]. Our findings validate this observation and raise questions about potentially defamatory content on these sites.

5.2 Data Access Requests

In this phase authors sent data access requests to the sites in the dataset in an attempt to access their own data on the site. While in the initial self-search we started observing connections between sites, in this phase, we discovered new ones. We explain them as relevant in this section and summarize the nature of the connections in Table 1.

5.2.1 Usability of Data Access Mechanisms.

We found that eight of the sites did not contain any automated method to request access to our data. In these cases we sent emails to the addresses provided or through their contact forms. For 10 other sites, through privacy policies, we found online forms that allowed users to exercise their privacy rights. AnyWho and 411 redirected to entire or parts of *fee* sites privacy policies (Intelius and Whitepages respectively).

Online forms were focused on data associated with user-created accounts on the site or were unsuccessful. Most of the online forms required multiple types of PII, such as physical and email addresses. In the case of Spokeo and PeopleFinders, the outcome of the automated forms was not what we expected. In PeopleFinders, the response of the form indicated that the site was unable to verify our identity or we encountered issues with the form submission. Completing the data access form on Spokeo's website resulted in a response indicating that "No data is associated with your account". However, we observed that in the most recent updated privacy policy at the time of the writing, the Right-to-Know request form in Spokeo can only be accessed when logged in, resulting in a detailed report of the user account information, without mention of public data access. Similarly, Intelius's user data tools, which indicated that it gives users the ability to exercise their "Right to Know", resulted in one of the authors receiving a file containing the searches they conducted using the provided email address. These findings indicate that these sites differentiate between user account data and user personal information, complicating the process of exercising data access rights. Standing out from other *fee* sites, PeekYou offers an online form (that contains an option to access information, different from the removal option). After submitting the form, we did not hear back.

In most of these cases, at least two of the researchers, being mindful of the processing load required by the sites customer's support staff, followed up with an email to the site's support team. We categorize and explain outcomes received in these cases along those received from the rest of the sites in the rest of this section.

Data access mechanisms revealed additional connections between sites. Emailing PeopleLooker resulted in a response by BeenVerified, the same one obtained through filling the form on the website of the latter; in Figure 2 we combine these sites together. For Intelius, Instant Checkmate, TruthFinder and USSearch we observed that sites contained two sets of forms "User Data Tools" and "Public Data Tools" with the same interfaces. We tried using both sites and while the first one led to different URLs for all of them, the latter lead to PeopleConnect, the parent company of all these sites. In Figure 2, we also combine these sites and AnyWho, previously found to link at Intelius's privacy policy. For the same reason, we combine Whitepages and 411.

5.2.2 Access Request Outcomes. We organize this section to mostly follow the themes in Figure 2.

We received actual reports only from group of sites. For three of the researchers, we managed to obtain a complete report from BeenVerified, after providing additional information to the customer support's staff. The reports were identical to the one we later accessed through purchase. The fourth researcher (non-CA/EU) did not manage to receive a report. The response indicated that "Identity couldn't be verified". This might have been due to a mismatch between the addresses provided for the verification process and the address that the site had. In the paid reports phase we find that BeenVerified includes an address in a state the author lived temporarily more than six years ago and did not provide as part of the identity verification step. While verifying the identity is a good way to ensure that data access requests are not being abused by adversaries, it is not known how the sites may use this information and if there are controls set in place to make sure it doesn't end up in future reports.

One site explained they are working on an automated process. Two of the researchers that sent follow-ups to Intelius received responses explaining that they are working on an automated method of enabling data access, but no alternate solution was offered.

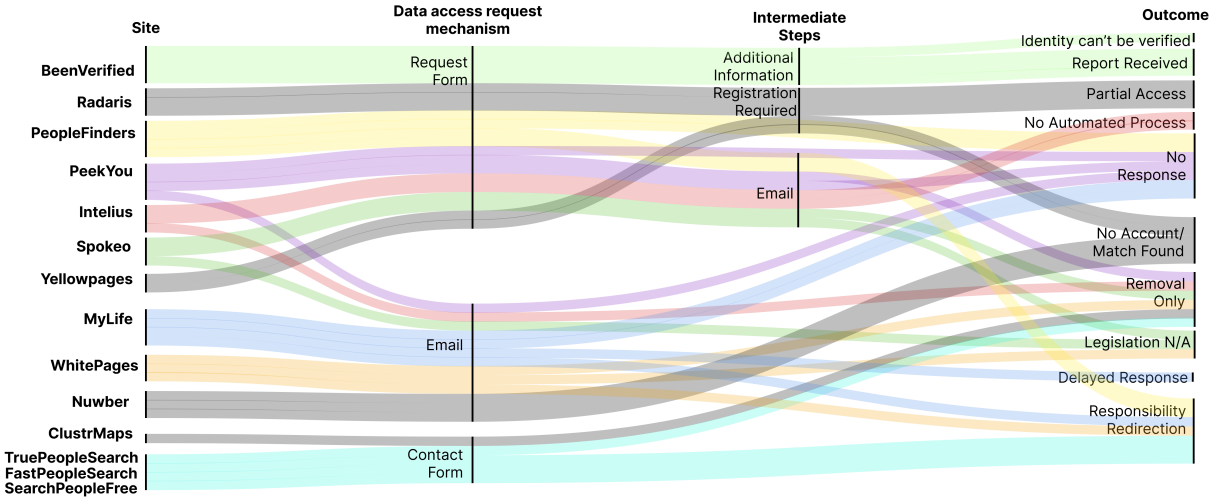

Figure 2: Visualization of the different interaction paths of the data access requests: We observed that responses received across different authors in the same sites were inconsistent with different reasons provided for data access rejections. Highlighted in different colors are the sites where we received inconsistent responses. Sites where we received consistent results for all authors who were listed are in grey. The thickness of the lines is proportional to the number of researchers who were listed on a site.

A few sites provided partial access. Radaris privacy policy claims: "If for any reason you are concerned about the personal information in your account, you have the ability to review your personal information online by logging into your account." Through a link in the footer, named "Control your Information", Radaris required creating an account to remove or edit personal information. As part of creating an account, a phone number and email was required. Our attempt to use VoIP numbers (Skype, Google Voice) was unsuccessful, so we left it up to individual researchers to choose to provide their phone numbers or not. We observed that even though advertised as a data access mechanism, the information obtained after creating an account is the same as the one that can be accessed without. We note that during the self-search phase, we observed that during the search process MyLife also has a way to indicate if you are the person you are searching, and like Radaris it requires registration.

While responses from Intelius indicated that they are working on an automated process for data access, we observed that the "Public Data Tools" form in PeopleConnect, even though designed to provide removal, could be used to gain partial access to data that this group of sites contained. Researchers observed that after providing date of birth and legal name they could select the profile that belonged to them. However, the data visible was mostly redacted, i.e. containing only the first six digits of the phone number.

Multiple sites reported no account or match was found.

After completing a form on the website, both researchers who were listed in Yellow Pages, had to create an account. After doing so and a few back and forths with the site's team, they received replies indicating that they "were unable to locate any registered consumer accounts associated with the email you provided". In this response, they also indicated that to remove listings from their other websites and apps, the researcher should submit a removal request to Intelius or contact phone service providers.

Researchers that were listed received similar responses from Nuwber, indicating that the email or name could not be matched. However in the case when the author residing in CA submitted the data access request, Nuwber rejected the data access request explaining that according to the data they have the state (CA), probably inferred by the CCPA request, does not match the one shown on the website (researcher's prior address). This reveals a conflict with using this information in evaluating the legitimacy of privacy rights, as the information in these sites, as in this case, might be outdated.

Multiple sites responded to Data Access Requests by offering removal. Two of the authors sent a data access request via email to PeekYou after the form submission wasn't successful. One of them (CA resident) resulted in the site offering to remove our information, while the other received no response. It is difficult to tell if this was differential treatment due to privacy legislation or just an inconsistency caused by the lack of an established process.

Offering removal was a common technique even when sites provided another reason to reject data access requests. In Figure 2, we do not include all instances removal was suggested to avoid overlapping of categories, but instead limit to sites that only offered removal (and did not provide any additional reason). However, other sites also explicitly suggested or implicitly included instructions to removal mechanisms (i.e. Clustrmaps, Nuwber, Spokeo and Whitepages). In one case, the latter did not wait for our reply and instead removed one of the researcher's record automatically. It is unclear if these sites misinterpreted our request or used this

response as a way to avoid providing data access reports. Moreover, removal is not a replacement for data access, especially in the case someone is trying to find out what data an attacker might have obtained from the site. Removal does not help individuals who are concerned about their privacy to form complete risk mental models.

Two sites claimed that legislation was not applicable. Two of the sites, Spokeo and Whitepages, responded to emails from the EU researcher, claiming that the GDPR does not apply to their data collection practices. One of these sites builds the argument around public data: "Spokeo and its data providers obtain the information from publicly available profiles that you have authorized, or did authorize, the social network to make public. [..] As such, in this context, Spokeo is neither a controller nor a processor of personal data under the GDPR".

Whitepages gives multiple reasons why the GDPR does not apply and offers to remove the author's information instead. It explains that the company does not monitor an individual's behaviour in the European Union and only markets its services for use within the U.S. and Canada. A related, high profile case around the use of public data of EU residents by a U.S. company is the legislative response received by Clearview AI. Similar to People Search Websites, Clearview AI also used public data, scraped from online social media profiles, to train its face recognition algorithm. Multiple EU legislators have fined Clearview for a breach of GDPR associated with the unlawful collection of data.

In the case of the CA-based researchers, emailing Spokeo resulted in the following response: "We collect publicly available information from a variety of public sources, including phone books, social networks, marketing surveys, real estate listings, business websites, and other public sources ("Public Information"). These types of information constitute "publicly available information" under state privacy laws and are therefore exempt or outside the scope of those laws." This response indicates that Spokeo utilizes the limitations of the CCPA definition of personal information to reject data access requests. As described in Section 2.3, the CCPA definition of personal information does not include publicly available information. However, it is unclear if the legislators predicted that this aspect would impact individual's right to access their data.

Multiple sites used responsibility redirection strategies. We consider responsibility redirection, instances when sites recommended we search for ourselves or suggested using the form on the website.

Three of the sites of *ad* sites (who don't offer paid subscriptions) TruePeopleSearch, FastPeopleSearch, SearchPeopleFree, recommended us to search ourselves as a response to data access requests. For example one of the emails stated: "If you are a California resident and are looking to exercise your right to know pursuant to the California Consumer Privacy Act, you simply need to run a search on yourself and this will pull up the information we show on you. Since this is a free website, you will never be asked to pay for information on our website." While three authors received a version of this response via email, the fourth one who did these steps later, found that two of these sites had added a version of this response to their contact form when "How do I access my records?" is selected. This set of responses is problematic because it puts the burden in the individual and it assumes that they can find out how to use the search functionality of the site. Further, given that these sites are *ad sites*, individuals who attempt to search themselves may be exposed to ads and tracking. When the PeopleFinders form wasn't successful, two of the authors attempted to seek data access by sending the request via email and received a response that suggested using the original form, another form of responsibility redirection.

Lastly, one of the authors was notified from MyLife that sufficient credit to search for oneself was added to their existing profile, another form of responsibility redirection. At this point the researcher had not created a profile, but we suspect that providing email to see information in the self-search phase might have been used or stored in the same way as user account information. Another author received a response from MyLife about two months later indicating that no account with the email address could be found.

5.3 Paid Reports

We purchased 12 reports, from the *fee and hybrid sites* in Table 1. In this subsection we summarize our observations about services offered, types of information included and similar information.

Most sites offered multiple tiers of subscriptions and additional services. In most cases, subscribing to the site signs the user up for the most basic subscription and once the user has signed up, multiple updates are advertised. Radaris and PeopleFinders allow selecting between a Basic and Premium service upfront. All PeopleConnect sites (Intelius, TruthFinder, USSearch and Instant Checkmate) advertise premium access and add-ons such as unlimited access to phone, email records and pdf reports. While we find that the information these sites offer is largely the same, the prices for the add-ons vary across sites. Purchasing the email and phone numbers add-ons we find that they did not contain any information that could not be deduced from the personal reports, which already contained researcher's contact information.

Intelius and USSearch also offered access to Premium data. For both of these sites the premium data included criminal and traffic records, assets, finances, business profiles and licences. We found that the premium tier did not provide any significant additional information. However, we acknowledge that this might not necessarily be true for all individuals listed on the sites and results might vary. In addition, outcomes might be different in the cases where the search subject has criminal records and significant assets.

Another type of services we noticed being offered are identity and credit protection services. Some of sites that offer these services are PeopleFinders, BeenVerified, Intelius (identity & credit protection), TruthFinder, Instant Checkmate ("Dark Web" scan). However, we leave the evaluation of these services to future work.

Fee sites supplement paid reports with additional information of limited utility. Information included in the paid reports was similar to the types of information accessible as a non paid customer, with *fee sites* providing additional details. As visible in 1, most *fee sites* which displayed partial addresses in their free profiles, included complete ones in the paid reports. Further, paid reports

often include additional contact information such as phone numbers or emails. However, we observed that emails were most often incorrect.

Other types of information unique to the paid reports were: possible jobs, miscellaneous results of internet searches, information about cellphone carriers, extensive lists of potential relatives/neighbors, which we often found to be only partially correct. It is not clear what methodologies are used to determine relatives and associates. In fact, the term *associates*, which often includes family members, seems to be purposely vague. Reports from PeopleConnect sites also included lists of sex offenders near the locations mentioned in our reports. For one of the researchers, these covered 11 pages making up a significant portion (over 1/3rd) of their report. While there is no doubt that this information can be useful, it is of a different nature from the rest of the report and might not be what a site visitor expects to find.

Multiple reports contained career information. As far as we can tell, LinkedIn might have been the source of such datapoints. However, we observed that the information was not always up to date with our Linkedin profiles, indicating that scraping might have happened in the past. Family-owned vehicles was one of the datapoints that was unique to the BeenVerified group and made the researchers particularly uncomfortable, due the high potential that this information has to track and follow someone down in real life. Similarly, Spokeo included an estimated income figure, which researchers were unable to understand how it could have been calculated. It also included several broad categories of interests (i.e. Health & Fitness, Cooking etc), which according to the report were according to information based on household and collected by companies and events for marketing usage.

Groups of sites contained similar information. All the sites that in the data access phase, we found to be part of the PeopleConnect family (Intelius, USSearch, TruthFinder, Instant Checkmate) have the same information. For example these sites were the only ones that included the address for the most recent move (same month at the time of the writing) for one of the authors. However, sometimes they will change the way they display information, for example Instant Checkmate displays a timeline that combines all the information, including date of birth, social media account creation, first appearance of living address and job information Sites from PeopleConnect largely have the same user-interfaces (UIs) with differences only in the color scheme and logo. Similarly, PeopleLooker and BeenVerified include identical information in the paid reports.

Among the six other *fee* and *hybrid sites*, many share different pieces of information, but we did not observe any other identical reports. For example, MyLife and BeenVerified had only an old temporary address for one of the authors, and PeopleFinders and Number had another non recent address for the same author.

We also observed similarities between some of the—as we later find—connected *fee* and *ad* sites. For example TruePeopleSearch, FastPeopleSearch and SearchPeopleFree, provide free reports with the same addresses and phone numbers as PeopleFinders (*fee site*). PeopleFinders' paid report only provides limited extras such as telephone company or dates the subject was seen at the address. On the contrary, Yellow Pages and AnyWho, *ad sites* did not include all of the addresses that the paid report from Intelius did.

5.4 Removing our Information

The various connection between different sites, indicated that some of these sites partner with each other and might be even owned by the same parent companies. To investigate if removal of information from a site impacts removal in its partners, we decided to do the removal in two phases and observe the effects. First we removed information from four sites (Intelius, BeenVerified, PeopleFinders and Whitepages). After two weeks we assessed consequences in the remaining sites and then requested removal from the ones that still remained. The number of total sites we send at least one removal request to (both first and second removal round) is ten.

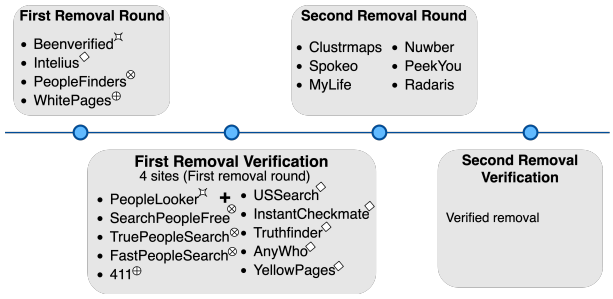

Figure 3: Visualization of the removal phase: In the first removal verification phase we discovered that upon requesting removal from four sites (first removal round) ten other sites had removed information. Symbols indicate the groups the sites belong to.

5.4.1 *Finding and Understanding Removal Mechanisms.*

Data removal mechanisms were hyperlinks in site footers. Seven out of ten sites we attempted removal from, included a version of "Do not Sell my Information" link in the website footer. One of these site's link redirected to the "Data Privacy Center" where a user can exercise multiple privacy rights. For the six other sites, the link redirected to an online form. Reading through the additional information this form included and comparing with the privacy policy instructions, we found that all but one of the forms was also meant to be used for profile removal requests. While it might be more efficient to use a single form to exercise multiple privacy requests, for a user seeking to remove their profile, it might not be obvious to click in the link that at first look, seems to prevent sale of information. The rest of the sites (3) included a hyperlink named "Remove (or Control) my info", more clearly indicating that these forms can be used to request removal from the site.

Information about removal scope and duration, as well as reappearance of information is incomplete.

To understand if and how these sites communicate relevant information about the data removal to users seeking to remove their information we looked at three main dimensions: scope, removal duration and information reappearance. The information about these topics was provided in the text accompanying the removal

forms. We found that scope was rarely well defined. For instance, Intelius removal procedure is the only one that explictly indicates that removal using this form, will also remove from other sites in the PeopleConnect family. After removing our information, four of the sites indicate that it might still be available from many other similar sites. Three of them link to paid removal services that might help with removal from multiple similar sites, such as OneRep (Nuwber, Clustrmaps) or BrandYourself (PeopleFinders). This observation comports with speculation in prior work [44], that there might be a symbiotic relationship between paid removal services and PSWs. The existence of PSWs introduces the need for paid removal services. On the other hand, paid removal services can only exist for as long as there is multiple PSWs and customers see the need to use a service instead of following a *do it yourself* approach.

Many sites explain how long the removal process will take. Most indicate either after form submission or in the confirmation email that the process will be completed in 48 hours or less. Only three of the sites do not give any indication of how long the removal will take. MyLife indicates inconsistent durations; after form submission that removal might take up to 15 days, while the confirmation email after form submission mentions 24 hours.

Reappearance of information has also been addressed. Six of the sites address reappearance of information, among which more than half of them (Peek You, Spokeo, BeenVerified, Nuwber) explain that the data might reappear if obtained from a different source. Two of the sites (Whitepages, BeenVerified) explain that they take measures to prevent reappearance, but only the latter explicitly mentions continuing to save user information in order to prevent a new listing being created in a future update. Therefore it is not clear if the other sites also store information for this purpose. This finding is in line with prior work that highlighted the lack of transparency with how data deletion requests are handled or processed and the lack of clear assurance that the data has been deleted [33].

5.4.2 *Usability of Removal Mechanisms.*

Data removal requests often required additional information and/or verification. Nine out of ten sites use a method of removal request verification; two of them request phone number for the verification process (Radaris, Whitepages), six of them request an email address and one (Intelius) allows both. In the removal request form only two of the sites require the physical address (Clustrmaps and MyLife). However, only four of the sites explicitly state that the information provided as part of the request will not be used for sale or any other purpose.

Seven of the removal forms request an URL of the profile that we want removed. One of them, Whitepages stands out in particular because the removal steps reveal the unredacted PII of the removal target, providing a way to receive access to one's own data. While this is a good way to verify that the removal is being requested for the right profile, it could also be leveraged by attackers to gain access to a target's data.

Incorrect information led to issues with identity verification. Removing information from Intelius was slightly more complicated. The online form sought to verify the author's identity, using the contact information listed in the site, i.e. by receiving a message or email. Given that for the researchers listed the contact information on Intelius was not correct, it was impossible to verify the identity using one of these methods. We emailed customer support, which helped us with manual removal, after requesting a few additional datapoints (name, DOB and location). This finding is in line with prior work on removal from People Search Websites that indicated that online forms are not always effective and sometimes users seeking removal need to contact the sites directly [44]. In addition, this occurrence surfaces an incorrect policy resolution. While under both the CCPA and GDPR businesses are supposed to verify the identify of individuals seeking to exercise privacy rights, People Search Websites cannot assume that the information they have can be used as "ground-truth" for such verification. Further, this reveals that PeopleConnect sites consider the information they include in the paid reports, both non-user/public data when we sought access and personal information in the removal phase.

Similarly, one of the authors' removal request from MyLife was initially unsuccessful. While no reason was provided, we suspect that it might have been due to the address on the site (out of date) being in a different state from the address provided in the removal form (current). Re-submission with the old address solved the issue. This observation reveals the need for clearly explaining how the information provided in these forms will be used and therefore improving transparency of removal methods. From a usability perspective, an error message indicating why the removal did not go through could help the users better understand and navigate the removal process.

Social media aggregator sites avoid removal. PeekYou and Radaris are distinguishable from the rest of the sites, as they seem to aggregate links from a variety of different sources. PeekYou focuses on social media networks and online question and answering sites like Quora, collecting all the potential accounts that might be associated with the search target. This process results in many unrelated profiles with the same first name being associated to a search target. Radaris follows a similar practice, but does not limit the scope to social media, including a wider variety of search results. In both of these cases the profile is a mix of results associated with multiple people. When attempting submit the PeekYou removal form, the instructions explained that only profiles with a certain URL structure, dedicated to a single person can be removed. Removal was successful for one of the authors with a dedicated profile. However two of the authors, whose profiles are amalgamated with others with the same first name did not manage to get their profile removed. Researchers also encountered issues with Radaris. As explained in the data access phase, Radaris requires account creation (and providing a phone number) to request removal. This finding is concerning, because it indicates that individuals have to decide between violating their privacy by sharing their phone number or allowing the profile to be visible. However, after claiming the profile, we found out two choices: making the profile private or deleting specific records. Selecting the latter, we found that we can remove only at most six datapoints from the profile with the search results not being one of the options. While it is true that the search results in Radaris were parts of other third party sites, the constrained removal options limit individual's agency to control their own public profiles. Further, we discovered that removing specific

datapoints (i.e. relatives, address) does not remove the profile itself. Making the profile private makes the profile unaccesible though search in the site.

5.4.3 *Assessing Removal.*

Paywalls can make it difficult to gauge removal success. The practice of hiding all datapoints behind a paywall (as observed in the self-search phase) in PeopleLooker and BeenVerified caused difficulties in gauging the removal success. For two of the authors we got a "Sorry, we have 0 results" prompt upon reattempting to search, indicating that the information was likely removed. However, for the rest of the researchers (more common names) it was impossible to verify without paying for the service. A simple search directed us to results hidden after the paywall, making it difficult to find out if those results are the authors themselves or people with the same name.

Removing from some of the *fee sites*, removed information from connected *ad sites* too. The original plan was to remove the information from four of the sites and then conduct a round of searches to understand the impact of the removal. However due to the above-mentioned verification issues encountered with Intelius, for two out of the three researchers listed, there was a delay in completing that removal. The delay enabled us to observe that after removing from BeenVerified, PeopleFinders and Whitepages, information was also removed from at least five sites. Due to connections observed through the data access phase we suspect that removal from PeopleLooker, was a result of requesting removal from BeenVerified. For a similar reason, we think 411 removal was attributed to Whitepages removal. We think that removal from PeopleFinders resulted in removal from TruePeopleSearch, FastPeopleSearch and SearchPeopleFree.

Removal from Intelius, once succeeded, impacted also removal from USSearch, Instant Checkmate, TruthFinder, AnyWho and Yellow Pages. We were not expecting the latter due to different access mechanisms observed in the data access phase. For the rest of the sites in this group, this finding confirms the observation that they all are connected.

No reappearance was observed after two months. We conducted another self-search two months after the first removal phase. In the sites we managed to successfully remove the information from, we did not observe any re-appearance of information. One of the researchers re-attempted accessing the data from BeenVerified, the only group we were somewhat successful in the data access request phase. BeenVerified replied with a response citing the removal date and explaining they don't have the data anymore. Similarly, trying to access the profile on Intelius did not show any matching profiles.

6 DISCUSSION

In this section we discuss the implications of our findings for improving public policy, technical solutions and security and privacy advice.

6.1 Policy Implications

Expanded CCPA definition of "personal information". Our findings indicate that many of the sites do not provide access to researchers' information, even when researchers could see that their information was available on the sites. The response received by Spokeo indicated that this might be due to a recent change to the CCPA that does not consider publicly available information PII. This reveals a duality between PSWs advertising information as valuable to visitors, but considering it "public information" once individuals attempt to access it. Further, CCPA's limited definition leaves users without any recourse and unable to obtain any transparency on their personal information, albeit public data, that these sites contain.

A similar differential treatment is also observed in sites in the PeopleConnect group (Intelius, USSearch, TruthFinder, Instant Checkmate). These sites have separate tools for accessing user data and publicly available data, indicating again that publicly available information is not considered personal information. When the researchers were trying to remove their information, the PeopleConnect site required identify verification using information (phone number or email) that the site already possessed. This reveals an inconsistency in how the data the sites contain is considered. In the verification step, these sites consider their information user data, but in the request for access they do not.

Improved guidelines for GDPR/CCPA policy interpretations. Some of the Data Access requests sent by the researcher based in the European Union received responses indicating that some sites don't have to comply with any GDPR requests. One of the sites claims that is is not "neither a controller nor a processor" because the only data it has about EU subjects is collected from social networks with public profiles. Our results from the paid reports phase reveal that this is not true as the site also includes the authors home address, which is not available on any social networks. Another site claims that it "processing of PII is not related to an offer directed to EU individuals" and it "does not monitor an individual's behavior in the Union". We note that a similar reason — claiming to have no EU business or customers — has also been used by Clearview AI as a response to fines from the French Data Protection Authority (CNIL) [28]. We think this similarity is interesting because in both cases these companies use public data, sometimes scraped from social networks, to build their products. Further, we find that some sites use IP geo-blocking likely as an attempt to block EU visitors. Besides making it difficult for EU residents, who might be listed on these sites, to exercise their privacy rights, IP blocking does not address the root of the problem. An EU citizen's data might still be listed and accessible on these sites and people residing in the EU can use a VPN if they want to purchase PII sold on the site, such as PII connected to another EU citizen.

Free and simplified access to individuals' profiles. Similar to the Fair Credit Reporting Act, that allows consumers to obtain a free credit report once a year, we argue that individuals should be able to obtain copies of the information that People Search Websites have collected about them. This would enable consumers to create thorough threat models. For example, our finding about data

access requests being largely unsuccessful indicates that if a target of online harassment might not be able to find out retroactively the exact information that attackers might have purchased access to. Similarly, our findings indicate that one is not able to use data access requests to identify cases of wrong or misrepresented information.

Requiring notices and improving transparency of "information flows". Researchers participating in the study had not created any of the profiles in the People Search Websites and were not sure how the information was collected. This finding is additional evidence of the drawbacks of the privacy notices (as described by "notice and consent"). It is likely that authors had agreed to provide their data to some entity which in turn sold it to People Search Websites, further highlighting that currently notices fail to include information about all possible information flows.

Further, our findings indicate that the sites prompt users for multiple data points when trying to exercise privacy rights or simply conduct a search for other individuals, but it is unclear how such data might be used. We found that MyLife added one of the emails provided during self-search to the user's profile. This finding indicates that these sites must be required to be transparent about how they plan to use addresses and other types of PII collected from their users. This finding also contributes to the discourse around "notice and consent", providing evidence that when information is the product, companies might not be incentivized to provide thorough notices.

6.2 Technical Recommendations

Mechanisms for reporting noncompliance. Our findings about PSWs being largely non-compliant with data access requests highlights the need for additional oversight in privacy rights compliance. Due to the highly sensitive data these sites possess, we argue that consumers should be able to report such cases to appropriate authorities. For example, mechanisms similar to FTC's fraud or identity theft reporting could be set in place.

While data removal requests were more often successful than not, we found that Radaris required creating an account to remove ones' profile. If implemented, the reporting mechanisms could also account for these cases. Given that users' consent was not obtained before listing information on these sites, we recommend that removal forms be accessible without requiring registration.

Reporting systems for misleading advertisements. We argue that consumers should be able to report misleading advertisements. For example, during the self-search phase, we observed that Spokeo states "15 court search results for people named [author's name] in the United States", but we found that paying for the background reports reveals that the author does not have any court records. It is not clear if these court records exist at all or are associated with another individual with the same name in the United States. These misleading indicators could be harmful to individual's reputation when a third party, even a non-malicious one, searches for them. This effect could disproportionately affect underrepresented groups. For example, prior work by Sweeney has found that PSW ads suggestive of arrest records differ by race [43]; with more of them using the word arrest appeared for black-identifying first names.

Further, exaggerated claims about what might be included in information hidden after a paywall are a way to incite curiosity. Even if the searcher wasn't intending to buy paid access, after being exposed to these advertising techniques they might, violating the target's privacy.

6.3 Security and Privacy Advice

Removal prioritization. Systematically mapping connections between the sites has broader implications for discourse around the "privacy paradox". One of the explanations of the privacy paradox suggests that while individuals may care about privacy, companies monetizing their data make privacy self-management difficult [41]. The existence of many PSWs can make removing one's information seem like an impossible task and cause users to "surrender" (i.e., believe that privacy self-management is impossible). The uncovered connections between fee and add sites provide evidence of the opposite - individuals can amplify their privacy-protecting efforts by strategically choosing sites for removal efforts that indirectly influence others.

Prior work by Take et al. [44] alludes to an almost symbiotic relationship between PSWs and paid removal services; the existence of these paid removal services relies the existence of many individual PSWs. If individuals perceive that removing their own information (for free) from each one is too difficult, they are more likely to use a service. In this work we find that some PSWs link to removal services (BrandYourself, One Rep) and provide actionable insights for simplifying privacy self-management (e.g., for those who cannot afford to hire a paid service).

7 CONCLUSION

In this work we conducted a multi-step systematic analysis of user privacy rights in 20 People Search Websites, sites that collect, catalog and sell individuals' personal information. First, we provided a characterization of the ecosystem of People Search Websites, identifying three different types (*fee*, *ad* and *hybrid*) based on their monetization strategies. We also identified connections between the sites, linking 14 sites to four distinct groups. Second, we investigated compliance with data access requests and found that most sites do not honor data access requests. Only one group of connected sites provided access to the same report that they provide to paying customers. Lastly, using a two-step experimental process, we found that removal from a set of four initial *fee sites* results in removal from connected *ad sites*, indicating that mapping connections between sites is an effective way to simplify removal.

ACKNOWLEDGMENTS

This manuscript benefited from excellent feedback from PoPETS reviewers and our anonymous shepherd. This work was supported by the National Science Foundation (grant number 2016061) and NOVA LINCS (ref. UIDB/04516/2020 and ref. UIDP/04516/2020) with the financial support of FCT.IP.

REFERENCES

[1] 2018. California Consumer Privacy Act of 2018 [1798.100 - 1798.199.100]. https://leginfo.legislature.ca.gov/faces/codes_displaySection.xhtml?lawCode=CIV&sectionNum=1798.110.

[2] 2018. California Consumer Privacy Act of 2018 [1798.100 - 1798.199.100]. https://leginfo.legislature.ca.gov/faces/codes_displaySection.xhtml?lawCode=CIV&sectionNum=1798.105.

[3] 2018. The Vermont Statutes Online: Title 9, Commerce and Trade, Chapter 062: Protection of Personal Information, 9 V.S.A. § 2446. https://legislature.vermont.gov/statutes/section/09/062/02447

[4] 2019. California Civil Code, Data Broker Registration [1798.99.80 - 1798.99.88]. https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?division=3.&part=4.&lawCode=CIV&title=1.81.48.

[5] No Date. General Data Protection Regulation (GDPR), Right of access by the data subject. https://gdpr.eu/article-15-right-of-access/

[6] No Date. General Data Protection Regulation (GDPR), Right to erasure ('right to be forgotten'). https://gdpr.eu/article-17-right-to-be-forgotten/

[7] Max Aliapoulios, Kejsi Take, Prashanth Ramakrishna, Daniel Borkan, Beth Goldberg, Jeffrey Sorensen, Anna Turner, Rachel Greenstadt, Tobias Lauinger, and Damon McCoy. 2021. A Large-Scale Characterization of Online Incitements to Harassment across Platforms. In *Proceedings of the 21st ACM Internet Measurement Conference*. Association for Computing Machinery, New York, NY, USA, 621–638. https://doi.org/10.1145/3487552.3487852

[8] Daly Barnett. 2020. Doxxing: Tips To Protect Yourself Online & How to Minimize Harm. https://www.eff.org/am/deeplinks/2020/12/doxxing-tips-protect-yourself-online-how-minimize-harm. Accessed: 2023-02-09.

[9] Luca Bufalieri, Massimo La Morgia, Alessandro Mei, and Julinda Stefa. 2020. GDPR: when the right to access personal data becomes a threat. In *2020 IEEE International Conference on Web Services (ICWS)*. IEEE, 75–83.

[10] Xavier Caddle, Ashwaq Alsoubai, Afsaneh Razi, Seunghyun Kim, Shiza Ali, Gianluca Stringhini, Munmun De Choudhury, and Pamela Wisniewski. 2021. Instagram data donation: A case for partnering with social media platforms to protect adolescents online. In *ACM Conference on Human Factors in Computing Systems (CHI 2021)/Social Media as a Design and Research Site in HCI: Mapping Out Opportunities and Envisioning Future Uses Workshop*.

[11] Heewon Chang. 2016. *Autoethnography as method*. Vol. 1. Routledge.

[12] Federal Trade Commission et al. 2014. Data brokers: A call for transparency and accountability. *Washington, DC* (2014).

[13] John Cook. 2010. Intelius quietly buys US Search, the 'godfather' of records search. https://www.bizjournals.com/seattle/blog/techflash/2010/04/intelius_quietly_buys_longtime_rival_us_search.html.

[14] James Denvil and Arielle Brown. 2021. CPRA countdown: Changes to the definition of "personal information". https://www.engage.hoganlovells.com/knowledgeservices/news/countdown-to-the-california-privacy-rights-act-changes-to-the-definition-of-personal-information. Accessed: 2023-10-03.

[15] David M Douglas. 2016. Doxing: a conceptual analysis. *Ethics and information technology* 18, 3 (2016), 199–210.

[16] Janna Lynn Dupree, Richard Devries, Daniel M. Berry, and Edward Lank. 2016. Privacy Personas: Clustering Users via Attitudes and Behaviors toward Security Practices. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 5228–5239. https://doi.org/10.1145/2858036.2858214

[17] Kevin Gallagher, Sameer Patil, Brendan Dolan-Gavitt, Damon McCoy, and Nasir Memon. 2018. Peeling the Onion's User Experience Layer: Examining Naturalistic Use of the Tor Browser. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security* (Toronto, Canada) *(CCS '18)*. Association for Computing Machinery, New York, NY, USA, 1290–1305. https://doi.org/10.1145/3243734.3243803

[18] Yael Grauer. 2020. How to Delete Your Information From People-Search Sites. https://www.consumerreports.org/electronics/personal-information/how-to-delete-your-information-from-people-search-sites-a6926856917/.

[19] Hana Habib, Sarah Pearman, Jiamin Wang, Yixin Zou, Alessandro Acquisti, Lorrie Faith Cranor, Norman Sadeh, and Florian Schaub. 2020. "It's a Scavenger Hunt": Usability of Websites' Opt-Out and Data Deletion Choices. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376511

[20] Hana Habib, Yixin Zou, Aditi Jannu, Neha Sridhar, Chelse Swoopes, Alessandro Acquisti, Lorrie Faith Cranor, Norman Sadeh, and Florian Schaub. 2019. An Empirical Analysis of Data Deletion and {Opt-Out} Choices on 150 Websites. In *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*. 387–406.

[21] Joanne Kim. 2023. Data Brokers and the Sale of Americans' Mental Health Data. (2023).

[22] Kristen Kozinski and Neena Kapur. 2020. How to Dox Yourself on the Internet. https://open.nytimes.com/how-to-dox-yourself-on-the-internet-d2892b4c5954. Accessed: 2023-02-09.

[23] Jacob Leon Kröger, Jens Lindemann, and Dominik Herrmann. 2020. How Do App Vendors Respond to Subject Access Requests? A Longitudinal Privacy Study on IOS and Android Apps. In *Proceedings of the 15th International Conference on Availability, Reliability and Security* (Virtual Event, Ireland) *(ARES '20)*. Association for Computing Machinery, New York, NY, USA, Article 10, 10 pages. https://doi.org/10.1145/3407023.3407057

[24] A Krolik and K Hill. 2021. The Slander Industry. https://www.nytimes.com/interactive/2021/04/24/technology/online-slander-websites.html.

[25] Equality Labs. 2017. ANTI-DOXING GUIDE FOR ACTIVISTS FACING ATTACKS. https://equalitylabs.medium.com/anti-doxing-guide-for-activists-facing-attacks-from-the-alt-right-ec6c290f543c. Accessed: 2023-02-09.

[26] Rachel Lerman. 2015. Intelius acquires Classmates for $30M. https://www.seattletimes.com/business/technology/intelius-acquires-classmates-for-30m/.

[27] Yijing Liu, Yan Jia, Qingyin Tan, Zheli Liu, and Luyi Xing. 2022. How Are Your Zombie Accounts? Understanding Users' Practices and Expectations on Mobile App Account Deletion. In *31st USENIX Security Symposium (USENIX Security 22)*. USENIX Association, Boston, MA, 863–880. https://www.usenix.org/conference/usenixsecurity22/presentation/liu-yijing

[28] Natasha Lomas. 2023. Clearview fined again in France for failing to comply with privacy orders. https://techcrunch.com/2023/05/10/clearview-ai-another-cnil-gspr-fine/.

[29] Alison R. Murphy, Madhu C. Reddy, and Heng Xu. 2014. Privacy Practices in Collaborative Environments: A Study of Emergency Department Staff. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Baltimore, Maryland, USA) *(CSCW '14)*. Association for Computing Machinery, New York, NY, USA, 269–282. https://doi.org/10.1145/2531602.2531643

[30] Craig Newman and Jonathan Schenker. 2019. Part II: A Closer Look at the CCPA's Definition of "Personal Information". https://wp.nyu.edu/compliance_enforcement/2019/05/10/part-ii-a-closer-look-at-the-ccpas-definition-of-personal-information/. Accessed: 2023-10-03.

[31] Helen Nissenbaum. 2011. A contextual approach to privacy online. *Daedalus* 140, 4 (2011), 32–48.

[32] Victor Le Pochat, Tom Van Goethem, Samaneh Tajalizadehkhoob, Maciej Korczyński, and Wouter Joosen. 2018. Tranco: A research-oriented top sites ranking hardened against manipulation. *arXiv preprint arXiv:1806.01156* (2018).

[33] Kopo M Ramokapane and Awais Rashid. 2023. ExD: Explainable Deletion. *arXiv preprint arXiv:2308.13326* (2023).

[34] Eduard Rupp, Emmanuel Syrmoudis, and Jens Grossklags. 2022. Leave no data behind–empirical insights into data erasure from online services. *Proceedings on Privacy Enhancing Technologies* 3 (2022), 437–455.

[35] Brennan Schaffner, Neha A. Lingareddy, and Marshini Chetty. 2022. Understanding Account Deletion and Relevant Dark Patterns on Social Media. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 417 (nov 2022), 43 pages. https://doi.org/10.1145/3555142

[36] Justin Sherman. 2021. Data Brokers and Sensitive Data on US Individuals. *Duke University Sanford Cyber Policy Program* 9 (2021).

[37] Justin Sherman. 2023. Data Brokers and the Sale of Data on U.S. Military Personnel. (2023).

[38] Robert H Sloan and Richard Warner. 2014. Beyond notice and choice: Privacy, norms, and consent. *J. High Tech. L.* 14 (2014), 370.

[39] Peter Snyder, Periwinkle Doerfler, Chris Kanich, and Damon McCoy. 2017. Fifteen Minutes of Unwanted Fame: Detecting and Characterizing Doxing. In *Proceedings of the 2017 Internet Measurement Conference* (London, United Kingdom) *(IMC '17)*. Association for Computing Machinery, New York, NY, USA, 432–444. https://doi.org/10.1145/3131365.3131385

[40] Daniel J Solove. 2000. Privacy and power: Computer databases and metaphors for information privacy. *Stan. L. Rev.* 53 (2000), 1393.

[41] Daniel J Solove. 2021. The myth of the privacy paradox. *Geo. Wash. L. Rev.* 89 (2021), 1.

[42] Keith Spiller. 2016. Experiences of accessing CCTV data: The urban topologies of subject access requests. *Urban Studies* 53, 13 (2016), 2885–2900.

[43] Latanya Sweeney. 2013. Discrimination in online ad delivery. *Commun. ACM* 56, 5 (2013), 44–54.

[44] Kejsi Take, Kevin Gallagher, Andrea Forte, Damon McCoy, and Rachel Greenstadt. 2022. "it feels like whack-a-mole": User experiences of data removal from people search websites. *Proceedings on Privacy Enhancing Technologies* 3 (2022), 159–178.

[45] Kejsi Take, Victoria Zhong, Chris Geeng, Emmi Bevensee, Damon McCoy, and Rachel Greenstadt. 2024. Stoking the Flames: Understanding Escalation in an Online Harassment Community. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 176 (apr 2024), 23 pages. https://doi.org/10.1145/3641015

[46] Emily Tseng, Rosanna Bellini, Nora McDonald, Matan Danos, Rachel Greenstadt, Damon McCoy, Nicola Dell, and Thomas Ristenpart. 2020. The Tools and Tactics Used in Intimate Partner Surveillance: An Analysis of Online Infidelity Forums. In *29th USENIX Security Symposium (USENIX Security 20)*. USENIX Association, 1893–1909. https://www.usenix.org/conference/usenixsecurity20/presentation/tseng

[47] Tobias Urban, Dennis Tatang, Martin Degeling, Thorsten Holz, and Norbert Pohlmann. 2019. A Study on Subject Data Access in Online Advertising After the GDPR. In *Data Privacy Management, Cryptocurrencies and Blockchain Technology: ESORICS 2019 International Workshops, DPM 2019 and CBT 2019, Luxembourg, September 26–27, 2019, Proceedings* (Luxembourg, Luxembourg). Springer-Verlag, Berlin, Heidelberg, 61–79. https://doi.org/10.1007/978-3-030-31500-9_5

[48] W Gregory Voss. 2021. The CCPA and the GDPR are not the same: why you should understand both. *W. Gregory Voss,'The CCPA and the GDPR Are Not the Same: Why You Should Understand Both,'CPI Antitrust Chronicle* 1, 1 (2021), 7–12.

[49] Miranda Wei, Sunny Consolvo, Patrick Gage Kelley, Tadayoshi Kohno, Franziska Roesner, and Kurt Thomas. 2023. "There's so Much Responsibility on Users Right Now:" Expert Advice for Staying Safer From Hate and Harassment. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 190, 17 pages. https://doi.org/10.1145/3544548.3581229

[50] Miranda Wei, Madison Stamos, Sophie Veys, Nathan Reitinger, Justin Goodman, Margot Herman, Dorota Filipczuk, Ben Weinshel, Michelle L. Mazurek, and Blase Ur. 2020. What Twitter Knows: Characterizing Ad Targeting Practices, User Perceptions, and Ad Explanations Through Users' Own Twitter Data. In *29th USENIX Security Symposium (USENIX Security 20)*. USENIX Association, 145–162. https://www.usenix.org/conference/usenixsecurity20/presentation/wei

[51] Ben Wolford. No Date. What is GDPR, the EU's new Data Protection Law? https://gdpr.eu/what-is-gdpr/

[52] Savvas Zannettou, Olivia-Nemes Nemeth, Oshrat Ayalon, Angelica Goetzen, Krishna P Gummadi, Elissa M Redmiles, and Franziska Roesner. 2023. Leveraging rights of data subjects for social media analysis: Studying TikTok via data donations. *arXiv preprint arXiv:2301.04945* (2023).

APPENDIX A

Due to the constantly changing nature of these sites, we provide a description below of the connections we observed at the time of data collection that informed our data removal process. In the Data Removal phase of the study, we utilized connections observed in the Data Access phase to determine which sites to remove from in the first removal round. In Table 2 below, we summarize the observed connections.

Removal Phase I	Suspected Connected Sites	Connection
Intelius	US Search Instant Checkmate Truthfinder	Similar information (Paid Reports Phase)
	AnyWho	Subject to Intelius privacy policy
Beenverified	PeopleLooker	Shared data access mechanisms. (Data Access Phase)
PeopleFinders	FastPeopleSearch TruePeopleSearch FastPeopleSearch	Ads contain affiliate marketing and lead to PeopleFinders
Whitepages	411	Privacy policy links to Whitepages

Table 2: Suspected Connections between sites that helped us divide sites into two removal phases.